

EUROPEAN PATENT APPLICATION

Application number: 82305926.6

Date of filing: 08.11.82

Int. CL²: **C 12 N 15/00**
C 12 N 1/00, C 12 P 21/02
C 07 H 21/04, C 07 C 103/52
//C12R1/19, C12R1/865

Priority: 12.11.81 US 320632

Date of publication of application:
25.05.83 Bulletin 83/21

Designated Contracting States:
BE CH DE FR GB IT LI NL SE

Applicant: **THE UPJOHN COMPANY**
301 Henrietta Street
Kalamazoo, Michigan 49001(US)

Inventor: **Dugalczyk, Achilles**
c/o The Upjohn Company 301 Henrietta Street
Kalamazoo Michigan 49001(US)

Representative: **Perry, Robert Edward et al.**
GILL JENNINGS & EVERY 53-64 Chancery Lane
London WC2A 1HN(GB)

Albumin-based nucleotides, their replication and use, and plasmids for use therein.

The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesized in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α -fetoprotein, but the synthesis decreases drastically after birth. Recently, Law et al determined the complete sequence of mouse α -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been reached from studies on the α -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched
 5 albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating
 10 the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending
 15 into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by
 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a
 25 prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table
 30 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

35 (5')...T^CT C T T C T G T.....albumin mRNA
 (3')...G A G G A A G G C G U C C m₂⁶A m₂⁶A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

231 val ser lys leu val thr asp leu thr lys val his thr glu oys oys his gly asp leu leu glu oys ala asp asp arg ala asp leu
 GTT TCC AGC TTA GTC ACA GAT CTT ACC AAA GTC CAC ACC GAA TGC TGC CAT GCA GAT CTT GAT CAC AGC GCG GAC CTT (890) 5 260

 261 ala lys tyr lle oys glu aan glin asp ser lle ser ser lys leu lys oys glu lys pro leu leu glu lys ser his oys lle
 CCC AAG TAT ATC TGT GAA AAT GAA CAT TGC ATC TCC AGT AAA CTC AAG GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TGC CAC TGT ATT (980) 289 290

 291 ala glu val glu aan asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val oys lys aan tyr ala
 CCC GAA GTC GAA AAT GAT GAG ATG CTT GCT CAC TTG CTT TTA GCT GCT CAT TTT GTT CAC AAG GAT GTT TGT AAA AAC TAT GTT (1070) 320

 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg his pro asp tyr ser val val leu leu leu arg leu ala
 GAG GCA AAG GAT GTC TTC TGC GGC ATG TTT TTG TAT GAA TAT GCA AAG CAT CCA TCT CTC GTC CTC CTC GTC ACA CTT GCG (1160) 350

 331 lys thr tyr glu thr thr leu glu lys oys oys ala ala asp pro his glu oys tyr ala lys val phe asp glu phe lys pro leu
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TGC TGT GCG CCA GAT CCT CAT GAA TGC TAT GCG AAA GAA GTC TTC CAT GAA TTT AAA CCT CCT (1250) 360 370

 381 val glu glu pro glin aan leu lle lys glin aan oys glu leu phe glu olin leu oly glu tyr lys phe olin aan ala leu leu val arg
 GTC GAA GAG CCT GAG AAT TTA ATC AAA GAA CAA AAT TGT GAG CTT TTT GAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTC TTA GTT CCT (1340) 410

 411 tyr thr lys lys val pro glin val ser thr pro thr leu val glu val ser arg aan leu oly lys val oly ser lys oys oys lys his
 TAC ACC AAG AAA GTA CCC GAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA ACA AAG CTA GCA AAA GTC GGC AAG AAA TGT TGT AAA TAT (1430) 437 438 440

 441 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu aan glin leu oys val leu his glu lys thr pro val ser
 CCT GAA GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC AAC CAG TTA TGT GTC TTG CAT GAG AAA AGC CCA GTA ACT (1520) 460 461 470

 471 asp arg val thr lys oys oys thr glu ser leu val aan arg arg pro oys phe ser ala leu glu val asp glu thr tyr val pro lys
 GAG ACA GTC ACC AAA TGC TGC ACA GAA TCC TTG GTC AAC AGC CCA CCA TGC TTT TCT CTC GTC CAT GAT GAA ACA TAC GTT CCC AAA (1610) 490 500

 501 glu phe aan ala glu thr phe thr phe his ala asp lle oys thr leu ser glu lys arg olin lle lys lys olin thr ala leu val
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GAT ATA TGC ACA CTT TCT GAG AAG GAG ACA ATC AAG AAA CAA ACT GCA CTT GCT (1700) 516 520 530

Following are examples which illustrate procedures, ~~including the best mode~~, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685].

15 Example 2 Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczky, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Bolívar, F., Rodríguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Royer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczky, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [³²P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30 As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris-HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.R., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

20 Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and γ [³²P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

30 Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(PHA36) - NRRL B-12551

5 HB101(PHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public, ~~upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.~~

10

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

15

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YE6 is a well known and widely available yeast episomal plasmid. It can be obtained from the following host deposit by standard procedures:

20

E. coli HB101 (YE6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straightforward task of restriction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

25

30

The assembled DNA clone can be used to transform two types of cells:

- 35
- (a) Escherichia coli
 - (b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

- (b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRI DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco RI cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.R. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

- The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Pujalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

25 Example 8 Screening of Clones Producing Albumin

- Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies
5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T.
10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- 5 2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
3. E. coli HB101 (pHA36) having the deposit accession number
10 NRRL B-12551.
4. E. coli HB101 (pHA206) having the deposit accession number
NRRL B-12550.
5. A microorganism modified to contain a nucleotide sequence
15 coding for the amino acid sequence of human serum albumin; said
nucleotide sequence is as follows:

20

25

30

35

231 val ser lys leu val thr asp leu thr lys val his thr glu oys oys his gly asp leu leu glu oys ala asp asp arg ala asp leu 260
 GTT TCC AGG TTA GCG ACA GAT CTT ACC AAA GTC CAC AGC GAA TGC TGC CAT GGT GAT GAC AGC GGC GAC CTT (890)
 261 ala lys tyr ile oys glu aan gln asp ser ile ser ser lys leu lys glu oys oys glu lys pro leu leu glu lys ser his oys ile 289 290
 GCC AGT TAT ATC TGT GAA AAT GAA GAT TGC ATC TCC AGT AAA CTC AGC GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TCC ATT (980)
 291 ala glu val glu aan asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val oys lys aan tyr ala 320
 GCC GAA CTC GAA AAT GAT GAG ATG CCT GCT GAC TTG CCT TCA TTA CCT GCT GAT TTT GTT GAA AGT AAG GAT GTT TGC AAA AAC TAT GCT (1070)
 321 glu ala lys asp phe leu glu met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu leu leu ala 350
 GAG GCA AAG GAT CTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA AGA AGC CAT CCT GAT TAC TCT GTC CTG CTG CTG ACA CTT GCC (1160)
 351 lys thr tyr glu thr leu glu lys oys oys ala ala asp pro his glu oys tyr ala lys val phe asp glu phe lys pro leu 380
 AAG ACA TAT GAA ACC ACT CTA GAG AGC TGC TGT GCT GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTG TTG CAT GAA TTT AAA CCT CTT (1250)
 381 val glu pro pro aan leu ile lys gln aan oys glu leu phe glu qln leu oly glu tyr lys phe gln aan ala leu leu val arg 410
 GTG GAA GAG CCT CAG AAT ATA ATC AAA CA AAT TGT GAG CTT TTT GAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CTT (1340)
 411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg aan leu oly lys val oly ser lys oys oys lys his 440
 TAC ACC AAG AAA GTA CCC GAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA AGA AAC CTA GCA AAA CTC GGC AGC AAA TGT TGT AAA CAT (1430)
 441 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu aan gln leu oys val leu his glu lys thr pro val ser 470
 CCT GAA GAA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AC CAG TTA TGT GTG TTG CAT GAG AAA ACC CCA GTA AGT (1520)
 471 asp arg val thr lys oys oys thr glu ser leu val aan arg arg pro oys phe ser ala leu glu val asp glu thr tyr val pro lys 500
 GAC AGA GTC ACC AAA TGC TGC ACA GAA TCC TTG GTG AAC AGC GCA CCA TCC TTT TCA CCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)
 501 glu phe aan ala glu thr phe thr phe his ala asp ile oys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val 530
 GAG TTT AAT CCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AGC GAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

| | | | |
|----|---|-----|-----|
| 5 | 1 | 10 | 20 |
| | asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asp phe lys | | |
| | ggt gca cac agc agt cag gtt cgt cgt ttt aaa gat ttg gca gaa gaa aat ttc aaa | | |
| 10 | 21 | 30 | 40 |
| | ala leu val leu lle ala phe ala gln tyr leu gln gln oys pro phe glu asp his val lys leu val asp glu val thr glu phe ala | | |
| | gcc ttg gtg ttg att gcc ttt gct cag tat ctt cag cag tct cca ttt gaa gat gta aaa tta gtg aat gaa gta act caa ttt gca | | |
| 15 | 51 | 60 | 70 |
| | lys thr oys val ala asp glu ser ala glu asp oys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu | | |
| | aaa aca tgt gtt gct gat cag tca gct gaa aat tct gac aaa tca ctt cat acc ctt ttt cca gac aaa tta tcc aca gtt gca act ctt | | |
| 20 | 81 | 90 | 100 |
| | arg glu thr tyr gly glu met ala asp oys oys ala lys gln glu pro gly arg asp oys phe leu gln his lys asp asp asp pro | | |
| | cgt gaa acc tat ggt gaa atg gct gac tcc tct cca aaa gaa gaa cca cct ggc aca aat gaa tcc ttg caa cac aaa gat aac aca | | |
| 25 | 111 | 120 | 130 |
| | asn leu pro arg leu val arg pro glu val asp val met oys thr ala phe his asp asp glu oys thr phe leu lys lys tyr leu try | | |
| | aac ctg ccc cca ttg gtg aca cca cgc ctt gat ctg atg tcc act gct ttt cat cac aat gaa gag aca ttt ttg aaa aac tac tta tat | | |
| 30 | 141 | 150 | 160 |
| | ala lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ala lys arg tyr lys ala ala phe thr glu oys gln | | |
| | gaa att gcc aca aca gat cct tac ttt tat gcc ccg gaa ctg ctt ttt gcc aat aac tat aaa gct gct ttt aca gaa tgt tcc caa | | |
| 35 | 171 | 180 | 190 |
| | ala ala asp lys ala ala oys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ala lys gln arg leu lys oys | | |
| | cct gct gat aaa gct gcc tcc ctg ttg cca aac ctg gat gaa ctt cgc gat gaa ggc agc gct tcc tct gcc aaa cag aca ctg aac tgt | | |
| | 201 | 210 | 220 |
| | ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu arg phe pro lys ala glu phe ala ala | | |
| | gcc agt ctg caa aaa ttt gca gaa aca gct ttc aaa gca tgg gca gta gct cgc ctg acc acc aca ttt ccc aaa gct gtt gca gaa | | |

231 val ser lys leu val thr asp leu thr lys val his thr glu cys oys his gty asp leu leu glu cys ala asp asp arg ala asp leu
 gtt tcc aag tta gtc aca cat ctt acc aaa gtc aac aac tgc tgc cat gca cat ctt gaa tct gct gat cac agc cgc gac ctt (1890)
 233 250 253 260
 261 ala lys tyr ile cys glu aan glin asp ser ile ser ser lys leu lys cys oys glu lys pro leu leu glu lys ser his oys ile
 gcc aag tat atc tgc gaa aat caa gat tgc atc tcc agt aaa ctc aag gaa tgc tgc gaa aaa cct ctc ttc gaa aaa tct cac tgc att (1980)
 265 270 278 279 280 289 290
 291 ala glu val glu asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asp tyr ala
 gcc gaa gtc gaa aat gat gag atg cct gct gac ttc tca tta gct gct att ttt gtt gaa agt aag cat ctt tgc aaa aac tat gct (1070)
 300 310 316 320
 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu leu
 gag gca aag gat gtc ttc ttc ggc atg ttt ttc tat gaa tat gca aga aag cat cct gat tac tct gtc ctc ctc gtc aca ctt gcc (1160)
 330 340 349 350
 351 lys thr tyr glu thr thr leu glu lys cys ala ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu
 aag aca tat gaa acc act cta gag aag tgc tgc tgc gct gca gat cct cat gaa tgc tat gcc aaa gtc ttc gat gaa ttt aaa cct cct (1250)
 360 361 369 370 380
 381 val glu pro glin aan leu ile lys glin aan oys glu leu phe glu leu tyr lys phe glin aan ala leu leu val arg
 gtc gaa gag cct gag aat tta atc aaa caa aat tgc gag ctt ttt gag cag ctt gca gag tac aaa ttc cag aat cgc ctc tta gtt cgt (1340)
 390 392 400 410
 411 thr lys lys val pro glin val ser thr pro thr leu val glu val ser arg aan leu gty lys val gty ser lys oys oys his
 tac acc aag aaa gta ccc caa gtc tca act cca act ctt gta gag gtc tca aca aac cta gaa aaa ctc gac aac aaa tct tot aaa cat (1430)
 420 430 437 438 440
 441 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val leu aan glin leu oys val leu his olu lys thr pro val ser
 cct gca gca aaa aca atg ccc tgc gca gaa gac tat cta tcc gtc ctc aac cag tta tot gtc ttc gat gag aaa acc cca gta agt (1520)
 448 450 460 461 470
 471 asp arg val thr lys oys oys thr glu ser leu val aan arg arg pro oys phe ser ala leu glu val asp olu thr tyr val pro lys
 gag aca gtc acc aaa tgc tgc aca gaa tcc ttc gtc aac aag cca cca tgc ttt tca cct ctc gaa gtc gat gaa aca tac gtt ccc aaa (1610)
 476 477 480 490 500
 501 glu phe aan ala glu thr phe thr phe his ala asp ile oys thr leu ser glu lys glu arg olin ile lys lys olin thr ala leu val
 gag ttt att gct gaa aca ttc acc ttc cat gca gat aia tgc aca ctt tct gag aag gag aca aac atc aag aaa caa cat gca ctt gtt (1700)
 510 514 520 530

5
10
15
20
25
30
35

558 559 560
531
540
567
570
580
590

331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

glu teu val lys his lys pro lys ala thr lys glu gin leu lys ala val met asp phe ala ala phe val glu lys eys eys lys
GAG CTC GTG AAA CAC AAG CCC AAG CAG ACA AAG CAG CAA CTC AAA GCT GTT ATG CAT TTC GCT TTT GTA CAG AAG TGC TGC AAG (17990)
531
540
567
570
580
590
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

-18 p r o -10
Met lys trp val tlu phe lle ser leu phe leu phe ser
GCTTTTCTCTCTGTCACCCCAAGCCCTTTGGCA ATG AAG TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT ACC (30)

-1 -6 p r o -1
e tyr ser arg gly val phe arg arg
T TAT TCC AGG GGT GTC TTT CGT CGA

[illegible]

231 val ser lys leu val thr asp leu thr lys val his thr glu oys his gly asp leu glu oys ala asp arg ala asp leu 260
 GTT TCC AGC TTA GTG ACA GAT CTT GAA GTC ACA ACC GAA TGC TGC CAT GCA CTT GAT CTT GAA TGT GCT GAT CAC AGC GCG GAC CTT (1890)
 261 ala lys tyr lle oys glu asn gln asp ser lle ser lys leu lys glu oys ala oys pro leu leu plu lys ser his oys lle 269 290
 GCG AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TGC TGT GAA AAA CCT CTG TTG CAA TCC CAC TGC ATT (980)
 291 glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val oys lys asn tyr ala 320
 GCG GAA GTG GAA AAT GAT CAG ATG CCT GCT GAC TTG CCT TCA TTA GCT GCT GAT TTT GTT GAA AGT AAG GAT CTT TGC AAA AAC TAT GCT (1070)
 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg his pro asp tyr ser val val leu leu leu arg leu ala 350
 GAG GCA AAG GAT GTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA AAT GCA AGC CAT CCT GAT TAC TCT GTC CTG CTG ACA TTT GCC (1160)
 351 tyr thr tyr glu thr thr leu glu lys oys ala ala asp pro his glu oys tyr ala lys val phe asp glu phe lys pro leu 380
 AAG ACA TAT CAA ACC ACT CTA GAG AAG TGC TGT GCG GCT GCA GAT CCT CAT CAA TGC TAT GCC AAA CTG TTC GAT GAA TTT AAA CCT CCT (1250)
 381 glu glu pro gln asn leu lle lys gln asn oys glu leu phe glu gln leu oly glu tyr lys phe gln asn ala leu leu val arg 410
 CTC GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA CTT CGT (1340)
 411 thr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu oly lys val oly ser lys oys oys lys his 440
 TAC ACC AAG AAA GTA CCC GAA GTG TCA ACT CCA ACT CTT GTA GAG CTC TCA ACA AAT CTA GCA AAA CTG GCG ACC AAA TGT TGT AAA CAT (1430)
 441 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu asn gln leu oys val leu his glu lys thr pro val ser 470
 CCT GAA GCA AAA ACA ATC CCC TGT GCA GAA CAC TAT CTA TCC CTG CTC CAG CAC TTA TGT GTG TTG CAT GAG AAA AGC CCA GTA ACT (1520)
 471 asp arg val thr lys oys oys thr glu ser leu val asn arg arg pro oys phe ser ala leu glu val asp glu thr tyr val pro lys 500
 GAC ACA GTC ACC AAA TGC TCC ACA GAA TCC TTG GTG GTC AAG AGC CCA CCA TGC TTT TCA GCT CTG GAT GAT GAA ACA TAC CTT CCC AAA (1610)
 501 glu phe asn ala glu thr phe thr phe his ala asp lle oys thr leu ser glu lys glu arg ala ile lys lys gln thr ala leu val 530
 GAC TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TGC ACA CTT TCT CAG AAG AAG ACA AAT CAC AAA CAA CAA CTT GCT (1700)

5
10
15
20
25
30
35

531
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala phe val glu lys oys lys
CAG CTC GTG AAA CAC CAC MAG CCC MAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT CAT TTC GCT GCT TTT GTA CAG MAG TGC TGC MAG (1790)

540
550
558 559 560

561
ala asp asp lys glu thr oys phe ala glu glu qly lys leu val ala ala ser gln ala leu qly leu ter
CCT GAC GAT MAG GAG ACC TGC TTT GCC GAG CAG GGT AAA AAT CTT GCT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CACACATTAAAG (1883)

567
570
580

ter
CATCTCAGCCCTACCATGGAATAGAGAGAGAAAAGTATTCATCTGCTTTTCTTTTCTGTTGTTAAAGCCACACGCCCTGCTCTAAMAAACATAATTCTTTAA (2002)

ter
TCATTTTCCCTCTTTTCTGCTGCTTCATTTAATAAATAATGCAAGCAATCTAA,..... 20AA (2078)

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

```

5      -16      p r c      -10
      Met lys trp val tlu phe lle ser leu leu phe leu phe ser
      ATG AAC TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT ACC (30)

10      20
      ser ala tyr ser arg gty val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gty glu asn phe lys
      TCG GCT TAT TCC AGG GGT GTG TTT CCG GAT CAT CCA CAC AAG ACT GAG CTT GCT CAT CCG TTT AAA GAT TTG GCA GAA GAA AAT TTC AAA (170)

15      30
      ala leu val leu lle ala phe ala gln tyr leu gln cys pro phe glu asp his val lys leu val asn glu val thr olu phe ala
      GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)

20      40
      lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gty asp lys leu cys thr val ala thr leu
      AAA ACA TGT GTT GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT GCA GAC AAA TTA TCG ACA GTT CCA ACT CTT (350)

25      60
      arg glu thr tyr gty glu met ala asp cys cys ala lys gln glu pro gty arg asn olu cys phe leu olu his lys asp asp asn pro
      CGT GAA ACC TAT GGT GAA ATG GCT GAC TCG TGT GCA AAA CAA GAA CCT GCG ACA AAT GAA TCG TTC TCG CAA CAC AAA GAT GAC AAC CCA (440)

30      80
      asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn olu thr phe leu lys lys tyr leu try
      AAC CTC CCC GCA TTG GTG ACA CAT CCA CAG GAT GAT GTG ATG TCG ACT GCT TTT CAT GAC AAT GAA GAC ACA TTT TTG AAA AAA TAC TTA TAT (530)

35      100
      ala lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr olu cys gln
      GAA ATT GGC ACA ACA CAT CCA TAC TTT TAT GCG CCG GAA CTC CTT TTT GCT AAA AGC TAT AAA GCT GCT TTT ACA GAA TGT TCG CAA (620)

      120
      ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp olu gty lys ala ser ala lys gln arg leu lys cys
      171 GCT GAT AAA GCT GCC TCG CTG TTG CCA AAG CTC GAT GAA CTT CCG GAT GAA GCG AAG GCT TCG TCG GCC AAA CAG ACA CTC AAG TGT (710)

      140
      ala ser leu gln lys phe gty glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala olu phe ala olu
      201 GCG ACT CTC GAA AAA TTT GCA GAA ACA GCT TTT AAA GCA TCG GCA GTA GCT CCG CTG AGC ACA TTT CCC AAA GCT CAG TTT GCA CAA (800)

      160
      220
      230

```

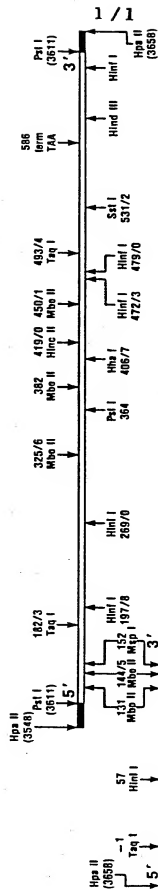

35

| | | | |
|--|-----|---------|-------------|
| 531 | 590 | 550 | 558 559 560 |
| glu leu val lys his lys pro lys ala thr lys glu gin leu lys ala val met asp asp phe ala phe val glu lys cys cys lys | | | |
| gag ctc ctg aac cag cag ccc aag gca acg aac gag cca ctg aac ggt gtt atg gat ttc gct gct ttt gta cag aag tgc aag | | | |
| 561 | 570 | 580 | |
| ala asp asp lys thr cys phe ala glu gln qly lys lys leu val ala ala ser gln ala ala leu qly leu ter | | | |
| gct gac gat aag cag acc tgc ttt gcc cag cag ggt aac aac ctt gtt gca agt cca gct tta gcc tta taa catcatttttaag | | | ter |
| | 567 | | |
| ter | | | |
| catc tgc tgc acc catc gatc atg agc aag aat gaa gca taa cgt ttt ttt ttc tgc tgc tgc tgc tgc aac gcc acc tgc tgc taa aac aat ttt ttt | | | (2002) |
| tcattttgccctcttttctctgcgtccattatataaaaatgcagagatctaa..... | 20 |aa | (2078) |

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13, 10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

Restriction Endonuclease Map of Human Serum Albumin cDNA Clones

pHA36



pHA206

